基于 SMOTE-NC 与 HRFT 混合模型的恶意流量多分类检测

赵嘉懿[1] 刘 铭[1*]
ZHAO Jiayi LIU Ming

摘 要 随着复杂网络攻击与流量激增对智能检测系统提出迫切需求。机器学习（machine learning, ML）和深度学习（deep learning, DL）技术在恶意流量检测领域展现出巨大潜力，但其在处理类别极不平衡的真实网络流量数据时，对少数类别攻击的检测性能往往不尽如人意。针对这一挑战，文章以广泛使用的 UNSW-NB15 数据集为基础，仅依赖网络流特征内容进行恶意流量的多分类研究。采用少数类过采样技术（SMOTE-NC）对训练数据进行增强。在增强后的数据集上重新训练 RandomForest 和 Transformer 模型，结果显示，尽管整体评估指标略有下降，但对少数类攻击的检测指标得到显著提升。基于此，进一步提出了一种新颖的并行混合模型 HRFT，该模型整合了 RandomForest 和 Transformer 的优势，并在经过 SMOTE-NC 增强的数据集上进行训练与测试。实验表明，混合模型在整体检测性能及对少数类攻击的识别能力上大部分优于 RandomForest 和 Transformer 这些基线模型，为构建更具有鲁棒性和均衡性的恶意流量检测系统提供了有效途径。

关键词 恶意流量检测；类别不平衡；SMOTE-NC；动态融合；注意力机制

doi：10.3969/j.issn.1672-9528.2025.11.029

0 引言

随着网络攻击的复杂化，传统网络安全防护手段在应对新型威胁时日益显得不足。基于机器学习（machine learning, ML）和深度学习（deep learning, DL）的智能入侵检测技术因其能自动从海量流量中学习攻击模式而备受关注[1]。然而，在实际应用中，尤其是在包含多种攻击类型且类别分布极不均衡的数据时，现有模型对少数类攻击的检测性能往往效果不佳[2]。

针对上述挑战，本文以 UNSW-NB15[3] 数据集为基础，首先采用 SMOTE-NC 技术增强训练数据以改善对少数类攻击的识别。随后，在增强数据集上重新评估 RandomForest 和 Transformer 模型的性能。最终，提出并验证了一种新颖的并行混合模型（HRFT），旨在整合 RandomForest 和 Transformer 的优势，以期在整体检测性能及对少数类攻击的识别能力上取得更优表现，为构建更鲁棒和均衡的恶意流量多分类检测系统提供有效途径。

1 相关工作

先前在网络入侵检测领域，已经存在很多相关研究探索。研究者们在 UNSW-NB15 和 KDD Cup '99 等数据集上评估了包括支持向量机（SVM）、决策树、随机森林（RF）、CNN、ANN 及结合了优化算法或注意力机制的朴素贝叶斯（PSO+NB）与长短期记忆网络（LSTM+Attention）等模型[4-9]。这些模型在二分类和多分类任务中表现各异，例如 Ahmad（2021）[7] 使用 RF 进行特征聚类选择在 UNSW-NB15 上取得了超过 97% 的多分类准确率。

2 算法与模型

2.1 数据集与预处理

本研究采用公开的 UNSW-NB15 数据集，该数据集包含多种现代网络攻击类型，并呈现显著的类别不平衡特性，适合验证本研究方法的有效性。预处理流程主要包括：

（1）特征选择：剔除原始数据中的 ID 列及用于二分类的 normal 标签，保留与多分类任务相关的 43 个原始特征。类别标签（如“DoS”等）转换为数值编码。

（2）数据标准化：对所有数值型特征采用 StandardScaler 进行标准化处理。

（3）类别型特征处理：对 proto、state 及 service 三个类别型特征进行独热编码（One-Hot Encoding）。处理过程中，为保证训练集、验证集与测试集特征维度统一，清除了极少量（<0.01%），因此产生维度不一致的样本。

2.2 SMOTE-NC 数据增强

针对 UNSW-NB15 训练数据中少数类攻击样本不足的

1. 北京警察学院 北京 102202

问题，本研究采用 SMOTE[10] 数据增强的变种 SMOTE-NC（SNC）对训练集进行数据增强。本研究中 SMOTE-NC 的实现关键步骤如下：

（1）以训练集中样本最多的类别（Normal 或 Generic）数量的一半或某个预设阈值为“目标样本数量”，将样本量少于此目标的攻击类别确定为待增强的少数类。

（2）其对数值型特征进行线性插值，对于一个少数类样本，SMOTE-NC 会找到其 k 个同属少数类的最近邻，随机选择其中一个，然后在这两个样本之间的连线上随机生成一个新的数值特征向量。用公式表示为：

$$x_{\text{new}}^{(i)} = x_i^{(j)} + \delta \cdot (x_{zi}^{(j)} - x_i^{(j)}) \quad (1)$$

式中：$x_i^{(j)}$表示原始少数类样本的第 j 个数值特征；$x_{zi}^{(j)}$表示其选定近邻的第 j 个数值特征；δ 表示 [0,1] 之间的随机数。

（3）对类别型特征进行基于其同类最近邻进行多数投票来确定新样本的特征值。用公式表示为：

$$x_{\text{new}}^{(F_{\text{cat}})} = \text{mode}\{x_j^{(F_{\text{cat}})} | x_j \in \{x_i\} \cup N_k(x_i)\} \quad (2)$$

式中：F_{cat} 表示任意类别特征；$N_k(x_i)$ 表示少数类样本 x_i 的 k 个少数类最近邻集合。

通过此策略，显著增加了训练集中稀有攻击类别的样本数量，为后续模型训练提供了更均衡的数据基础。

2.2.1 模型总体架构

HRFT 模型采用并行架构，如图 1 所示。输入的流量样本被同时送入 RandomForest 分支和 Transformer 分支。两分支独立进行特征学习后，其输出被送入一个动态融合层，该层负责自适应地整合来自两个分支的信息，并作出最终的分类决策。

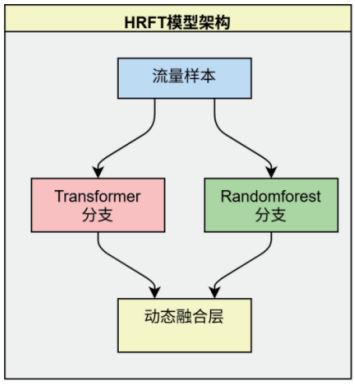

图 1 HRFT 模型整体架构图

2.2.2 RandomForest 分支

该分支通过构建多棵决策树并综合其结果，能够有效捕捉特征间的非线性关系，并对噪声数据具有较好的鲁棒性。在本模型中，RandomForest 分支输出每个类别的预测概率。

2.2.3 Transformer 分支

Transformer 分支的输入特征首先通过嵌入层映射到高维向量空间，随后经过多层 Transformer 编码器处理。该分支输出的是一个维度为 d_{model}（256 维）的特征向量，这个特征向量随后经过维度调整，与随机森林模型的预测概率进行融合。

2.2.4 融合层设计

为有效整合 RandomForest 和 Transformer 两个并行分支的信息，本研究探索了 3 种融合策略，其中动态注意力融合为本文主要提出的方法。

（1）动态注意力融合（DA）

详细结构如图 2 所示。

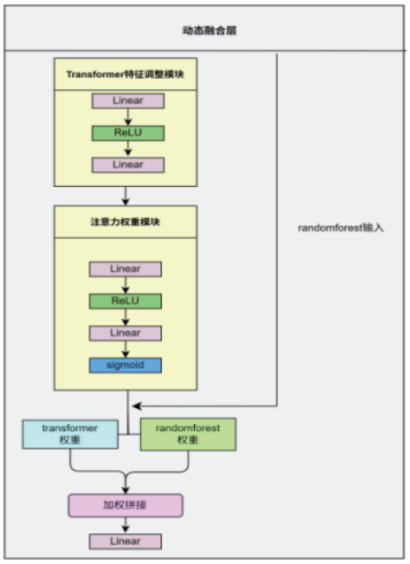

图 2 HRFT 模型动态融合层结构图

① Transformer 特征调整：Transformer 分支输出的 d_{model} 维特征向量首先经过一个小型前馈网络（包含线性层、ReLU 激活）进行维度调整和非线性变换，使其输出维度与任务的类别数量对齐，便于后续与 RandomForest 分支的输出（类别概率）进行有效融合。

② 注意力权重生成：调整后的 Transformer 特征被送入一个独立的注意力网络，这里的注意力权重模块为每个样本动态生成一个融合权重 $\alpha \in [0,1]$。此权重反映了注意力权重模块对，Transformer 分支的置信度。

③ 动态加权与融合：利用权重 α 对调整后的 Transformer 分支输出进行加权 $\alpha \cdot \text{output}_{\text{Transformer}}$，同时对 RandomForest 分支的输出（类别概率）进行互补加权 $(1-\alpha) \cdot \text{output}_{\text{RandomForest}}$，随后，将这两个加权后的特征向量进行融合，融合公式为：

$$\text{Fused} = \alpha \cdot \text{output}_{\text{Transformer}} + (1-\alpha) \cdot \text{output}_{\text{RandomForest}} \quad (3)$$

④ 最终决策：将融合特征送入一个最终的全连接（线性）层进行分类，输出最终的恶意流量类别预测。

该融合机制的核心优势在于其动态性和端到端可学习性。融合权重由模型根据输入数据自行学习得到，使得 HRFT 模型能够根据不同样本的特性，侧重于更可靠的分支信息。

（2）可学习加权平均融合（LWA）

此方法采用一个全局共享的可学习参数（经 Sigmoid 约束到 0~1 之间）来静态地平衡两分支的贡献。融合过程通过计算调整后的 Transformer 分支输出和 RandomForest 分支输出的加权平均值来，公式为：

$$\text{Fused} = \omega \cdot \text{output}_{\text{Transformer}} + (1-\omega) \cdot \text{output}_{\text{RandomForest}} \quad (4)$$

（3）直接拼接融合（DC）

调整后的 Transformer 分支输出与 RandomForest 分支输出不经过任何显式加权，直接沿特征维度拼接。这个拼接后的向量维度是原来的两倍，包含了两个分支的所有信息，随后被送入最终决策层。决策层在训练过程中隐式地学习如何组合和利用这两个信息源进行分类。

3 实验与结果分析

3.1 评价指标

准确率（Ac）：定义为被正确分类的样本数占总样本数的比例。公式为：

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (5)$$

式中：TP 表示真正例；TN 表示真负例；FP 表示假正例；FN 表示假负例。

精确率（Pr）：指在所有被模型预测为正类的样本中，实际也为正类的样本所占的比例。公式为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

召回率（Re）：指在所有实际为正类的样本中，被模型正确预测为正类的样本所占的比例。公式为：

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FP}} \quad (7)$$

F_1 值：精确率和召回率的调和平均数，用于综合评估模型的性能，尤其在处理类别不平衡的数据集时更为有效。

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision+Recall}} \quad (8)$$

3.2 SMOTE-NC 数据增强对比实验

为评估 SMOT-NC 的有效性，选用了 Transformer 与 RandomForest 模型，在 UNSW-NB15 数据集上 SMOTE-NC 数据增强前后进行对比。

从表 1 所体现的整体性能指标来看，SMOTE-NE 处理后，两个模型的准确率、召回率和 F_1 分数均略有下降，但精确率有小幅提升。这种现象可能是因为引入的合成样本可能带有一定的噪声，但这是为了换取对少数类攻击更优检测能力的权衡。

各类别 F_1 分数如表 2 所示，关键在于 SMOTE-NE 显著提升了对原先难以检测的少数类攻击的识别能力。对于原始数据中样本量极少的类别，RandomForest 和 Transformer 的 F_1 分数在使用 SMOTE-NE 增强数据后均得到大幅提升。

然而，对于 Shellcode 这一特定少数类别，其 F_1 分数在数据增强后出现了不升反降的情况 。这可能由于 Shellcode 攻击本身多样化和隐蔽性的特征，导致原始训练集中有限的 Shellcode 样本不足以充分代表该类别的整体分布；SMOTE-NC 的插值机制在面对特征极其复杂、非线性关系强或与其他类别边界高度重叠的少数类时，生成的合成样本质量可能受限，甚至可能引入一些不具代表性或模糊的样本，反而干扰了模型的学习。

3.3 HRFT 混合模型性能评估

在验证 SMOTE-NE 有效性的基础上，本研究进一步评估了所提出的 HRFT 混合模型在经过 SMOTE-NE 增强的数据集上的性能。

如表 1 所示，HRFT 模型在增强训练数据集上的准确率、召回率和 F_1 分数上均优于增强训练后的单一 RandomForest 及 Transformer 模型。特别地，HRFT 模型在召回率上展现出更佳表现，表明此模型在保持较低误报率的同时，能更有效地识别出恶意活动。

表 1 各模型整体性能

模型	SNC	Ac	Pr	Re	F_1
RandomForest		0.767	0.834	0.767	0.774
	√	0.712	0.847	0.712	0.759
Transformer		0.763	0.819	0.763	0.772
	√	0.674	0.847	0.674	0.730
HRFT-DA	√	0.715	0.842	0.715	0.760
HRFT-LWA	√	0.725	0.833	0.725	0.766
HRFT-DC	√	0.726	0.849	0.726	0.769

由表 2 可见，HRFT 模型在多数攻击类别及正常流量类别上，均达到或略微超过了两个增强后的基线模型中的最优者。

表 2 各模型的各类别 F_1 分数

	RandomForest		Transformer		HRFT-DA	HRFT-LWA	HRFT-DC
SNC		√		√	√	√	√
Normal	0.85	0.80	0.85	0.76	0.81	0.83	0.82
Generic	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Exploits	0.70	0.69	0.69	0.66	0.69	0.67	0.70
Fuzzers	0.40	0.38	0.37	0.35	0.37	0.37	0.41
DoS	0.11	0.33	0.12	0.29	0.32	0.30	0.33
Recon	0.85	0.81	0.84	0.80	0.80	0.82	0.81
Analysis	0.01	0.05	0.00	0.05	0.04	0.08	0.06
Backdoor	0.05	0.09	0.05	0.09	0.09	0.06	0.09
Shell	0.51	0.30	0.42	0.26	0.31	0.27	0.35
Worms	0.04	0.16	0.09	0.11	0.18	0.13	0.20

3.4 消融实验

为评估 HRFT 模型中不同融合层策略的有效性，本研究比较了前文的三种机制。所有实验均在 SMOTE-NE 增强后的数据集上进行。

实验结果表 1 显示，在整体性能指标上，三种机制存在一定差异。HRFT-DC（直接特征拼接）在准确率、召回率和 F_1 分数上均取得了最高的数值，而 HRFT-DA 则在精确率上略占优势，但其准确率和 F_1 分数稍逊于 DC 。HRFT-LWA 的各项整体指标则介于两者或略低于 DC 。

三种融合方法均显著优于单一模型在增强数据上的表现，但 LWA 策略相较于 DA 和 DC 整体表现稍逊。LWA 采用的单一全局融合权重难以适应网络流量数据中样本的高度异质性。

对比 DA 和 DC，尽管动态注意力机制（DA）在理论上具有更强的自适应学习能力，能够根据每个样本的特点动态分配权重，但在本次实验的各项指标上，其性能提升相对于信息保留更完整的直接特征拼接（DC）策略而言，优势并不显著，甚至在 F_1 分数等关键综合指标上略逊于 DC。从表 2 的各类别 F_1 分数来看，HRFT-DA 和 HRFT-DC 在不同攻击类别上的表现各有千秋，比较接近。可以推测，经过两个分支的特征本就具有高度的互补性与判别力，将其直接拼接，其效果可能不亚于甚至超过更复杂的动态加权机制。

4 结论

本文针对网络恶意流量检测中类别不平衡导致的少数类识别难题，提出了 SMOTE-NC 增强下的 HRFT 混合模型。实验表明，在 UNSW-NB15 数据集上，HRFT 相比单一增强模型，关键指标如 F_1 分数与召回率均有提升，并显著改善少数类攻击检测。同时，HRFT 通过动态融合 RF 与 Transformer 优势，兼顾多数类识别，展现了对整体流量更均衡鲁棒的检测能力，为构建高效入侵检测系统提供了有效途径。

参考文献：

[1] 蹇诗婕 , 卢志刚 , 杜丹 , 等 . 网络入侵检测技术综述 [J]. 信息安全学报 , 2020, 5(4):96-122.

[2]TALUKDER A, ISLAM M, UDDIN A, et al.Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction[J/OL].Journal of big data, 2024[2025-10-23]. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-00886-w.

[3]MOUSTAFA N, SLAY J . UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C/OL]// 2015 Military Communications and Information Systems Conference (MilCIS).Piscataway:IEEE,2015[2025-10-25].https://ieeexplore.ieee.org/document/7348942.DOI: 10.1109/MilCIS.2015.7348942.

[4]JING D S, CHEN H B. SVM based network intrusion detection for the UNSW-NB15 dataset[C/OL]// 2019 IEEE 13th International Conference on ASIC (ASICON). Piscataway:IEEE,2019[2025-10-21].https://ieeexplore.ieee.org/document/8983598.DOI：10.1109/ASICON47005.2019.8983598.

[5]KASONGO S M, SUN Y X. Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset[J/OL]. Journal of big data,2020[2025-10-25].https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00379-6.

[6]TALITA A S, NATAZA O S, RUSTAM Z. Naïve bayes classifier and particle swarm optimization feature selection method for classifying intrusion detection system dataset[J]. Journal of physics: conference series,2021,1752(1):12021.

[7]AHMAD M, RIAZ Q, ZEESHAN M,et al. Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set[J/OL].EURASIP journal of wireless communications and networking, 2021[2025-10-22].https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-021-01893-8.

[8] 张双全 , 殷中豪 , 张环 , 等 . 基于残差卷积神经网络的网络攻击检测技术研究 [J]. 信息网络安全 , 2025, 25(2): 240-248.

[9] 卢达 , 冯中华 . 基于卷积神经网络的网络入侵检测技术研究 [J]. 通信技术 , 2024, 57(5): 529-536.

[10]CHAWLA N V, BOWYER K W, HALL L O,et al.SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002,16: 321-357.

【作者简介】

赵嘉懿（2004—），男，北京人，本科，研究方向：人工智能与网络安全。

刘铭（1977—），通信作者 (email:liuming@bjpc.edu.cn)，男，辽宁辽阳人，硕士，教授，研究方向：网络安全与执法等。

（收稿日期：2025-05-21 修回日期：2025-11-12）